Agriculture
Canada

Research Branch
Direction générale de la recherche

Procedures for computing and mapping Thiessen weighting factors from digitized district boundaries and climatological station latitudes and longitudes

Canada

Procedures for computing and mapping Thiessen weighting factors from digitized district boundaries and climatological station latitudes and longitudes

G. D. V. WILLIAMS
Canadian Climate Center
Atmospheric Environment Service
Environment Canada
Downsview, Ontario
(formerly with Land Resource Research Institute)

H. N. HAYHOE
Agrometeorology Section
Land Resource Research Institute
Ottawa, Ontario

LRRI Contribution No. 82-26

Research Branch
Agriculture Canada
1982

Copies of this publication are available from:
Dr. H. N. Hayhoe
Agrometeorology Section
Land Resource Research Institute
Research Branch, Agriculture Canada
Ottawa, Ontario
K1A 0C6

Produced by Research Program Service

ABSTRACT

This paper describes and illustrates a system which incorporates a computerized grid mesh technique for determining the Thiessen weighting factors to estimate areal averages from point data. The set of procedures was developed in conjunction with research on the estimation of large area crop yields from weather and soils data. The system requires the latitudes and longitudes of climatological stations, and boundary coordinates obtained from digitizing a map showing the crop reporting districts. The package of computer programs includes transformations which reduce climatological station coordinates and digitized crop district boundaries to a common rectangular coordinate system. The effectiveness of the procedures is demonstrated numerically, and graphically through the use of the SYMAP computer mapping program, for crop reporting districts in the province of Saskatchewan. The application of the computed weighting factors is illustrated for the June 1978 rainfall.

RESUME

Le présent rapport décrit et illustre un système basé sur une technique de quadrillage informatisé permettant de déterminer les facteurs de pondération de Thiessen utilisés pour évaluer des moyennes superficielles à partir de données ponctuelles. Le procédé a été mis au point en conjonction avec une recherche relative à l'évaluation des rendements agricoles de grandes surfaces à partir de données sur le climat et les sols. Pour appliquer le système, if faut connaîtres les latitudes et les longitudes des stations climatologiques ainsi que les coordonnées des limites obtenues par conversion numérique d'unme carte des districts producteurs. La progiciel comprend des transformations qui réduisent les coordonnées des stations climatologiques et les valeurs numériques des limites des districts producteurs à une même système de coordonnées rectangulaires. L'efficacité du procédé est démontrée numériquement et graphiquement par l'intermédiaire du programme informatisé de cartographie SYMAP utilizé avec les districts producteurs de la province de Saskatchewan. L'application des facteurs de pondération calculés est illustrée pour la précipitation de juin 1978.

1. INTRODUCTION

The Thiessen polygon method for estimating areal data from point observations has several applications in meteorology and geography (Rhynsburger, 1973). The concept is quite simple, but adequate for many purposes: it assumes that the data value anywhere can be considered to be the same as that at the nearest observation point. If, for example, the monthly rainfall of a district is to be estimated for a number of months, and 65% of that district's area is closest to weather station A and 35% of station B, the formula used is $R_D = .65R_A + .35R_B$, where R_D is the district rainfall for a month and R_A and R_B are the station rainfall values for that month. However, in a computer analysis, the manual geometric construction and planimetric methods that are normally used to partition the district and compute the weighting factors can be quite inconvenient.

The authors have been involved in large area crop yield analyses (Williams, Joynt and McCormick, 1975; Sheppard and Williams, 1976) which required reducing weather and soils data to a crop district basis so that they could be related to the district yields. Sometimes for such work weather data that may be available on the basis of districts that at least approximately correspond to the crop districts can be used, as was the case in the United States spring wheat analyses of Starr and Kostrow (1978). In other cases the data may not be available on this basis, and even if they are there may be reasons for using some particular set of climatological stations, especially where operational yield prediction is involved.

In preparing the weather data in the regional yield analyses in which we were involved, the Thiessen polygon method was used to obtain district values from the point data (Williams and Robertson, 1965; Newton-Smith and Williams, 1964). Because of the work involved in manually re-determining weighting factors for different sets of stations, the set of precipitation stations that had been available during the initial phases of the work in 1962 continued to be used until 1975, even though this meant that full advantage was not being taken of the weather data that were later available. Problems also arose whenever there were closures or discontinuities in records among the fixed set of stations.

Further crop forecasting work (Williams and Lattimore, 1978) required the use of a different set of stations (Fig. 1) and the possibility of employing a program developed by Louie (1977) to compute the required Thiessen weighting factors was investigated. This program requires station and boundary locations to be provided in a rectangular Cartesian coordinate system, whereas station locations are normally given in terms of latitudes and longitudes. These can be converted to a rectangular system, but it is not convenient or desirable to convert them to the same system as used in digitizing the boundaries, and different parts of the boundary set may themselves be in different systems if they were digitized at different times. Another consideration here is that a strictly computational approach to determining Thiessen weighting factors generally would preclude the possibility of visual inspection which could otherwise be helpful in making a final selection of stations to be used.

We have therefore developed procedures (Fig. 2) which take area boundaries as digitized from a map, and the latitudes and longitudes of the stations to be used, and compute the required Thiessen weighting factors. The procedures also provide computer produced maps representing the polygons and weighting factors if desired. The system incorporates the Thiessen weighting factor program of Louie (1977), together with several other programs. The purpose of this paper is to describe this system or set of procedures and illustrate its application.

2. DESCRIPTION OF THE SYSTEM

a. Transformation of coordinates

The latitudes and longitudes of all climatological stations in Canada are readily available. In agrometeorological studies station data are often plotted for overlaying on maps using the Lambert Conformal Conic projection with standard parallels at 49° and 77°N. This projection gives good directional and shape relationships for a country such as Canada, which has a great east-west extent, and it is commonly used in maps showing all of Canada or large parts of it.

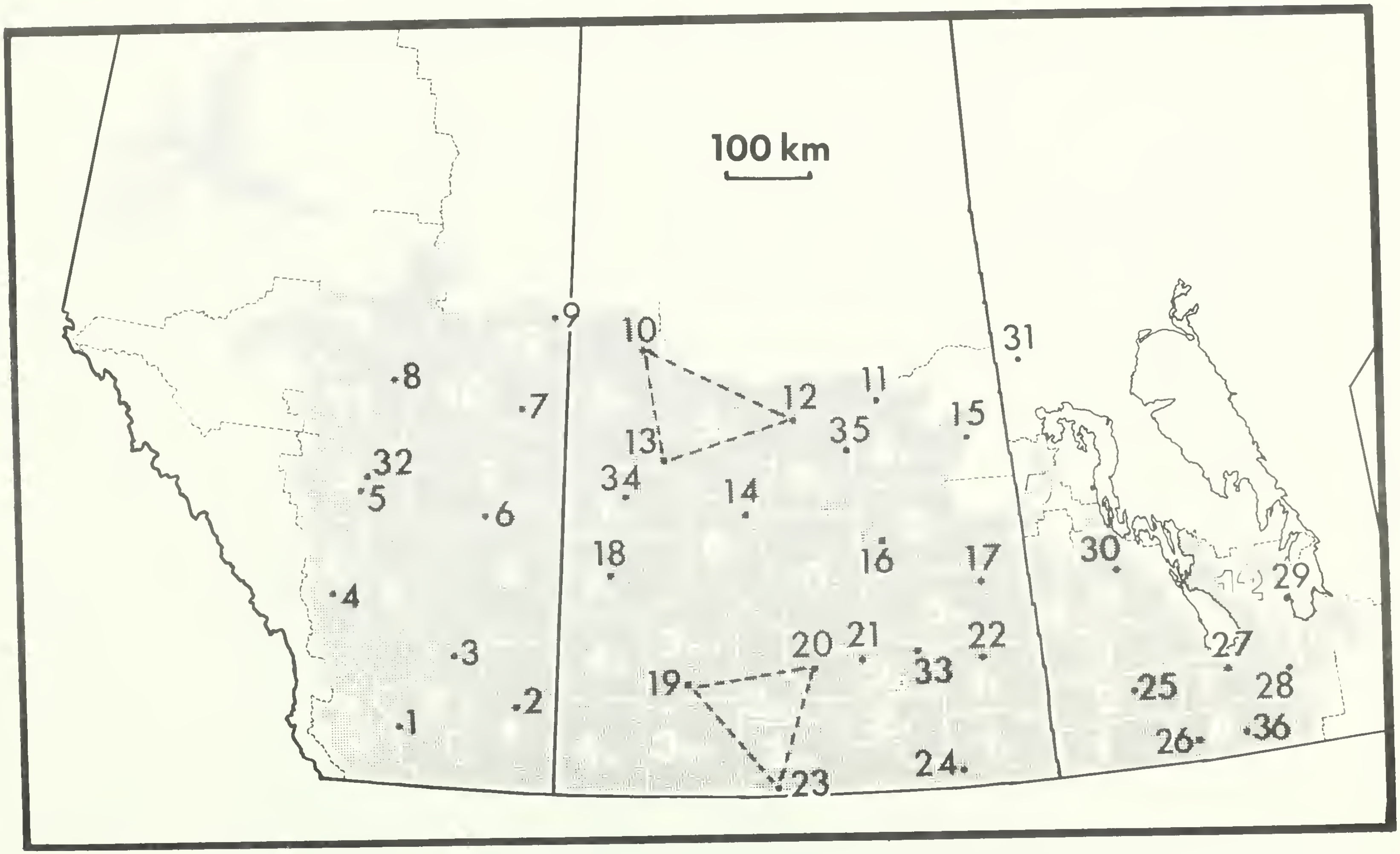


Fig. 1 Map of crop district boundaries in the provinces of Alberta, Saskatchewan and Manitoba, showing climatological stations listed in Table 1. The shaded area shows the approximate extent of the cereal zone.

In preparing the weather data in the regional yield analyses in which we were involved, the Thiessen polygon method was used to obtain district values from the point data (Williams and Robertson, 1965; Newton-Smith and Williams, 1964). Because of the work involved in manually re-determining weighting factors for different sets of stations, the set of precipitation stations that had been available during the initial phases of the work in 1962 continued to be used until 1975, even though this meant that full advantage was not being taken of the weather data that were later available. Problems also arose whenever there were closures or discontinuities in records among the fixed set of stations.

Further crop forecasting work (Williams and Lattimore, 1978) required the use of a different set of stations (Fig. 1) and the possibility of employing a program developed by Louie (1977) to compute the required Thiessen weighting factors was investigated. This program requires station and boundary locations to be provided in a rectangular Cartesian coordinate system, whereas station locations are normally given in terms of latitudes and longitudes. These can be converted to a rectangular system, but it is not convenient or desirable to convert them to the same system as used in digitizing the boundaries, and different parts of the boundary set may themselves be in different systems if they were digitized at different times. Another consideration here is that a strictly computational approach to determining Thiessen weighting factors generally would preclude the possibility of visual inspection which could otherwise be helpful in making a final selection of stations to be used.

We have therefore developed procedures (Fig. 2) which take area boundaries as digitized from a map, and the latitudes and longitudes of the stations to be used, and compute the required Thiessen weighting factors. The procedures also provide computer produced maps representing the polygons and weighting factors if desired. The system incorporates the Thiessen weighting factor program of Louie (1977), together with several other programs. The purpose of this paper is to describe this system or set of procedures and illustrate its application.

2. DESCRIPTION OF THE SYSTEM

a. Transformation of coordinates

The latitudes and longitudes of all climatological stations in Canada are readily available. In agrometeorological studies station data are often plotted for overlaying on maps using the Lambert Conformal Conic projection with standard parallels at 49° and 77°N. This projection gives good directional and shape relationships for a country such as Canada, which has a great east-west extent, and it is commonly used in maps showing all of Canada or large parts of it.

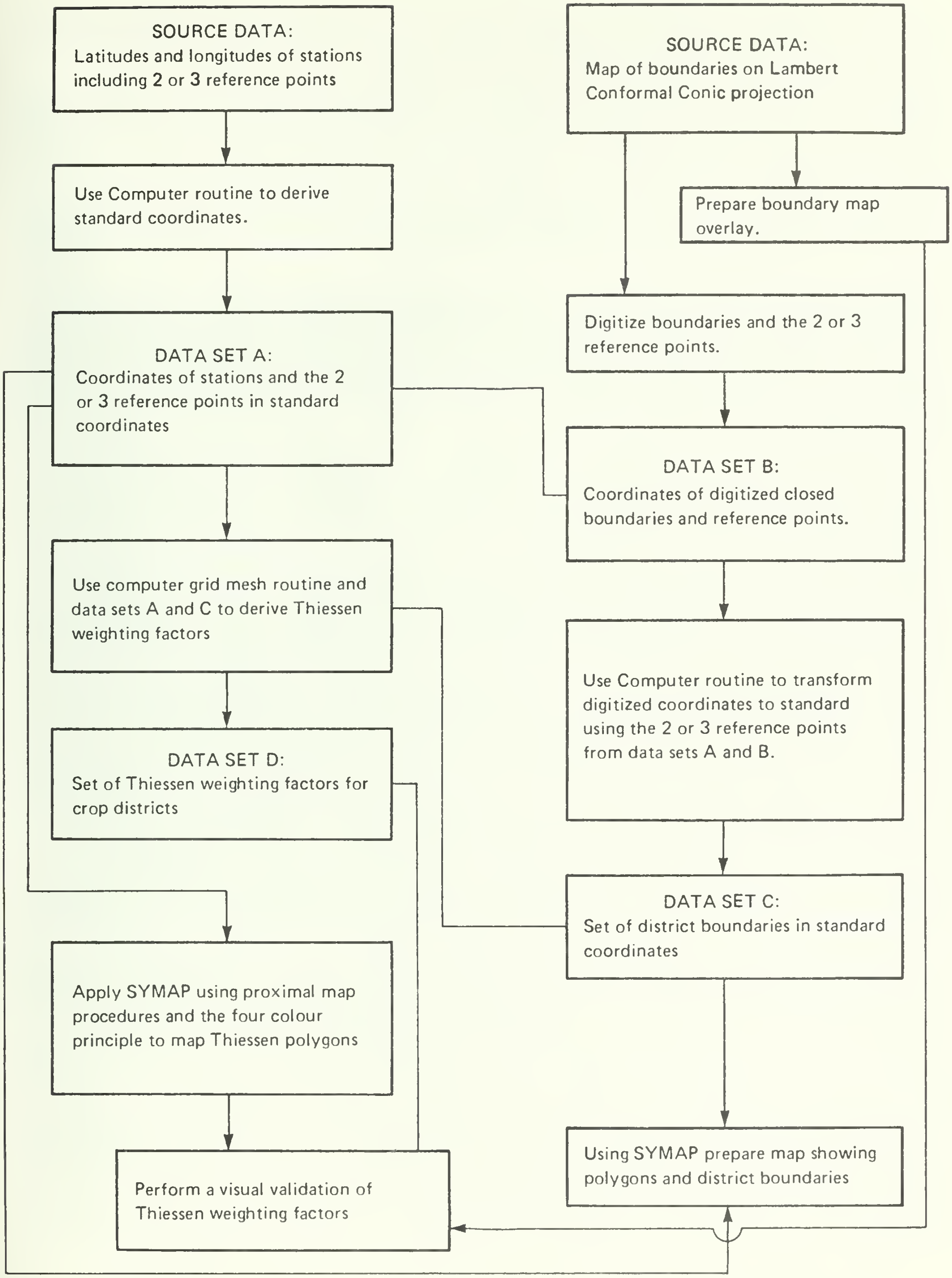


Fig. 2 Flow chart of the system used to calculate and display graphically the Thiessen weighting factors.

A computer subroutine obtained from the Canada Department of Energy, Mines and Resources is used to convert the geographical coordinates in degrees and minutes to X and Y values in a rectangular Cartesian coordinate system. The authors have adopted a system, to be referred to as the standard coordinate system throughout the remainder of this paper, in which the 95°W meridian of longitudes is vertical. The X value is the distance to the right, and Y the distance up, from an origin 533 mm to the left of the 41°N 95°W point on a 1:5,000,000 scale map. Thus Y measured toward the top edge of the map along the 95th meridian represents distance northward.

The basic requirements with regard to data for computing Thiessen weighting factors, or mapping the polygons by computer as described here, are a set of latitudes and longitudes of stations and a source map showing the district boundaries. There should be some reference points on the source map, preferably more than two, for which the latitudes and longitudes are known.

The district boundary map must be digitized to provide the boundary information for input to the computer if this has not already been done. This involves recording X and Y coordinates for points along each boundary, and also for each reference point. This is usually done on a digitizing table equipped to record the coordinates on tape or cards as the operator moves a cursor along the line being digitized, although for small jobs it can be done manually with the aid of graph paper. The units of measurement employed for the coordinates normally depend on the equipment used, and the positioning of the axes is determined by convenience except that we recommend placing the west edge of the map at the top for digitizing.

Using the same conversion routine as in the case of the climatological stations, the latitudes and longitudes of the reference points are converted to the standard coordinate system. The digitized boundary coordinate data are then transformed by linear transformation procedures to the standard coordinate system as used for the station locations.

The computer software to quantify the linear transformation and then to apply it to the digitized coordinate data is combined in one computer program. Firstly, the coordinates of the reference points in the standard and in the digitized coordinate system are read into the computer and the parameters of the linear transformation are established by solving a system of linear equations. The digitized coordinate data are then read and transformed to the standard coordinate system using the appropriate transformation matrix. We will describe in more detail how this is carried out for the case where three reference points are used and for the special case where only two are available.

Let XS(K), YS(K) denote the coordinates of the Kth point in the standard coordinate system and XD(K), YD(K) denote the coordinates of the same point in the digitized coordinate system. The linear transformation relating these two coordinates can be expressed in matrix notation as:

$$\begin{bmatrix} XS(K) \\ YS(K) \end{bmatrix} = \begin{bmatrix} A(1,1) & A(1,2) \\ A(2,1) & A(2,2) \end{bmatrix} \begin{bmatrix} XD(K) \\ YD(K) \end{bmatrix} + \begin{bmatrix} B(1) \\ B(2) \end{bmatrix} \quad (1)$$

We note that in (1) there are six unknowns, namely, the four A(I,J)'s and the two B(I)'s.

For the three-reference point case we denote the coordinates of the reference points in the standard coordinate system by (XRS(K), YRS(K)) K = 1,2,3 and in the digitizing system, by (XRD(K), YRD(K)) K = 1,2,3. From equation (1) the following two matrix equations are formed:

$$\begin{bmatrix} XRS(1) \\ XRS(2) \\ XRS(3) \end{bmatrix} = \begin{bmatrix} 1 & XRD(1) & YRD(1) \\ 1 & XRD(2) & YRD(2) \\ 1 & XRD(3) & YRD(3) \end{bmatrix} \begin{bmatrix} B(1) \\ A(1,1) \\ A(1,2) \end{bmatrix} \quad (2)$$

and

$$\begin{bmatrix} YRS(1) \\ YRS(2) \\ YRS(3) \end{bmatrix} = \begin{bmatrix} 1 & XRD(1) & YRD(1) \\ 1 & XRD(2) & YRD(2) \\ 1 & XRD(3) & YRD(3) \end{bmatrix} \begin{bmatrix} B(2) \\ A(2,1) \\ A(2,2) \end{bmatrix} \quad (3)$$

The computer system reads in the digitized and standard coordinates of the reference points and solves these two equations for the A(I,J)'s and the B(I)'s using Gaussian elimination with partial pivoting (Dahlquist et al., 1974). The digitized coordinate data are then read and transformed into the standard coordinate system using equation (1).

An analysis of the nature of the transformation required to relate the digitizing table coordinate system to the standard system suggests that there would normally be valid assumptions that could reduce the number of parameters to be estimated from six to four. This in turn reduces the number of reference points required from three or two. The assumptions are that the coordinate axes on the digitizing table can be transformed to the standard axes by performing a rigid rotation, a translation and an identical scale change in both the X and Y coordinates.

To motivate the restraints that these assumptions put on the matrix A in equation (1), consider the rigid rotation of two perpendicular axes through an angle O. Take a fixed point with coordinates, (X(1), Y(1)) in the initial coordinate system and (x(1), y(1)) in the new system. Using simple trigonometric relations we can specify the transformation with:

$$\begin{bmatrix} x(1) \\ y(1) \end{bmatrix} = \begin{bmatrix} \cos 0 & \sin 0 \\ -\sin 0 & \cos 0 \end{bmatrix} \begin{bmatrix} X(1) \\ Y(1) \end{bmatrix} \tag{4}$$

If a translation and a scale change identical in both directions is combined with the above rotation, we obtain the general relation

$$\begin{bmatrix} XS(K) \\ YS(K) \end{bmatrix} = \begin{bmatrix} A(1,1) & A(1,2) \\ -A(1,2) & A(1,1) \end{bmatrix} \begin{bmatrix} XD(K) \\ YD(K) \end{bmatrix} + \begin{bmatrix} B(1) \\ B(2) \end{bmatrix} \tag{5}$$

Comparing (1) with (5) we note that, as in (4), A(2,1) = -A(1,2) and A(2,2) = A(1,1). For this system only four parameters require estimation to specify the transformation. To evaluate the four parameters our computer system reads in (XRS(K), YRS(K)) and (XRD(K), YRD(K)) for K = 1,2. It then forms from equation (5) the matrix equation

$$\begin{bmatrix} XRS(1) \\ XRS(2) \\ YRS(1) \\ YRS(2) \end{bmatrix} = \begin{bmatrix} 1 & 0 & XRD(1) & YRD(1) \\ 1 & 0 & XRD(2) & YRD(2) \\ 0 & 1 & YRD(1) & -XRD(1) \\ 0 & 1 & YRD(2) & -XRD(2) \end{bmatrix} \begin{bmatrix} B(1) \\ B(2) \\ A(1,1) \\ A(1,2) \end{bmatrix} \tag{6}$$

Equation (6) is solved for A(1,1), A(1,2), B(1) and B(2) using the same Gaussian elimination routine. The digitized coordinate date then can be transformed into the standard coordinate system using equation (5).

b. Weighting factor computations

After the locations of the district boundaries and meteorological stations have been transformed to a standard rectangular coordinate system, they are ready for use as input to the program to calculate the Thiessen weighting factors.

Rather than working analytically with polygons, we adopted the grid mesh technique developed by Louie (1977). This technique requires that the boundary coordinates for a district be read in a clockwise order and that all the station coordinates be entered. The computer program then establishes a rectangular grid which encloses the district. Grid points lying within the district are systematically checked to see which station they are closest to. The proportion of points in the district closest to a station determines that station's Thiessen weighting factor.

Required levels of accuracy are achieved by iteratively refining the grid mesh. Louie (1977) tested the method and provided an illustration of the one-district case.

Our type of application involves a large region which is divided into number of districts, with stations scattered across the region (Fig. 1). Our initial approach was to only enter the stations with each district that could reasonably be expected to have a non-zero Thiessen weighting factor. After testing the procedure where all the stations were included with each district it was found that the routine was capable of quickly eliminating every station with a zero Thiessen coefficient. Our system is therefore designed to read in the coordinates of all the available stations. It then reads in the digitized boundary and performs the Thiessen calculations for each district in turn.

The number of iterations per district required by the grid mesh technique is controlled by the rate of convergence and the level of accuracy specified by the user. In addition, we specify that there should be a minimum of four and a maximum of seven iterations. The reason for using a minimum of four iterations is that we observed cases where, for the coarse grids used in the first few iterations, a station could be assigned a zero Thiessen weight, but the weight subsequently became non-zero.

When the grid mesh technique was applied in preliminary testing to the 42 crop districts and 65 precipitation stations that had been used in previous analyses, the computer Thiessen weighting factors corresponded quite well to those that had been determined earlier by much more laborious methods.

c. Visual representation

The standard coordinates of the stations and boundaries prepared for use in the computation of the Thiessen weighting factors can readily be employed as input for the SYMAP computer mapping program (Dougenik and Sheehan, 1976). This program utilizes the standard computer printer to prepare maps, using combinations of printed and overprinted characters to produce patterns. Other applications of this program in agrometeorology have been demonstrated by Williams and Sharp (1972) and by Williams, McKenzie and Sheppard (1980).

The proximal map option of SYMAP is well adapted to mapping Thiessen polygons. With it one can use different patterns for different polygons, and the edges between polygons may be left as white lines, or they can be lines of printed symbols. Alternatively, the polygons can be left white and only their outlines shown (Fig. 3).

The boundaries of the districts can be depicted at the same time by using the "Otolegend" package of SYMAP (Fig. 4). Thus a

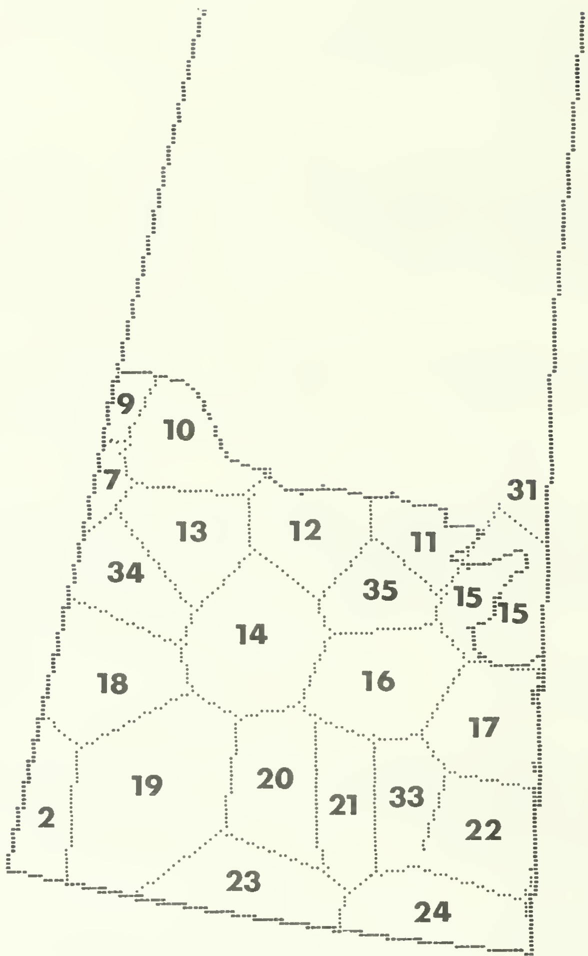


Fig. 3 Map of Saskatchewan crop growing area prepared with SYMAP, showing Thiessen polygons for climatological stations numbered as in Table 1.

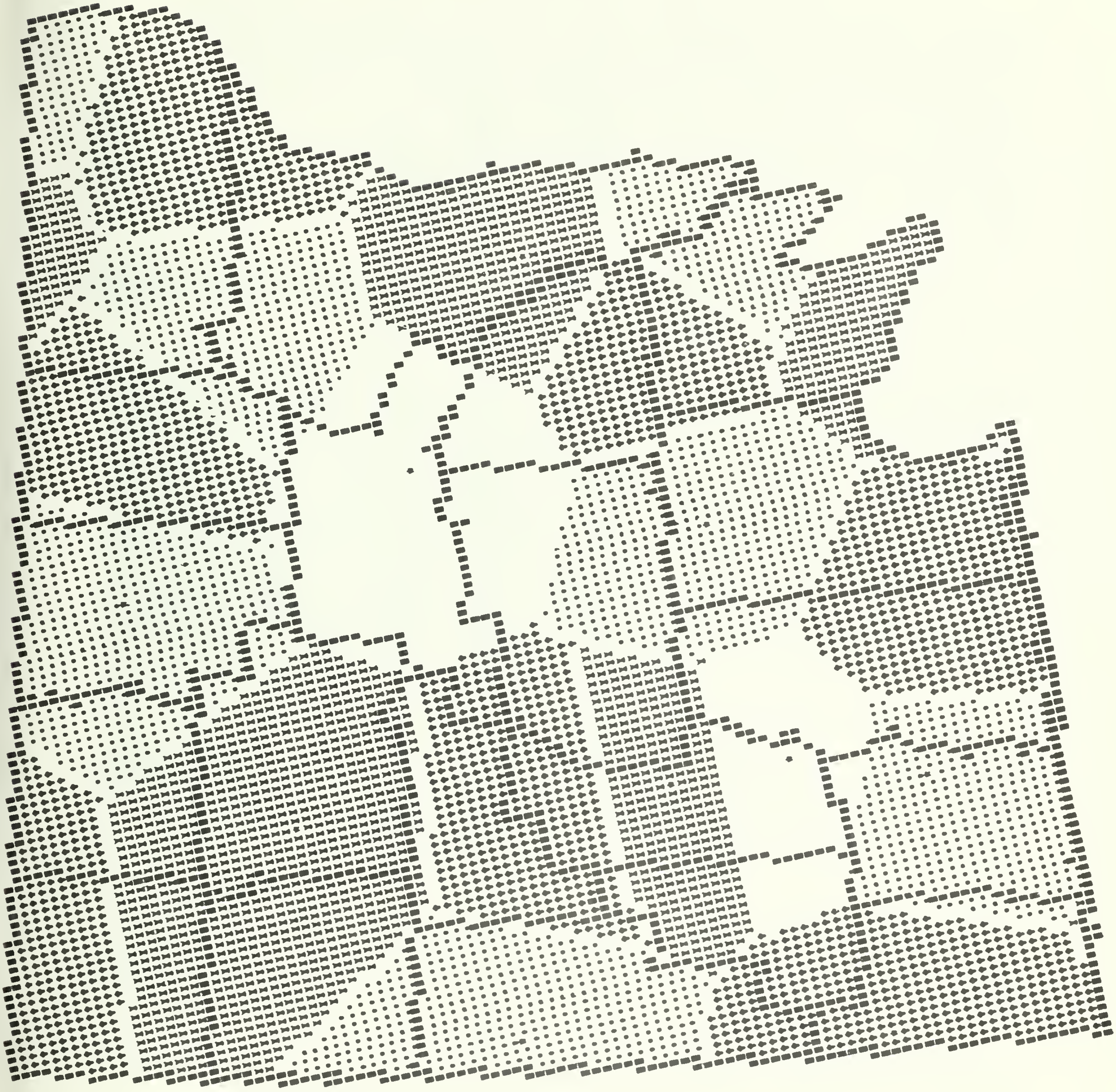


Fig. 4 SYMAP map showing Saskatchewan crop districts overlaid on the Thiessen polygons.

visual representation can be provided of the proportions of each district that are in various polygons, i.e. the computer Thiessen weighting factors are represented by relative areas on the map that is printed out using SYMAP.

In mapping with colours one could use only four colours to prepare a map of polygons such that no edge between adjacent polygons had the same colour on both sides. Similarly, a "four-level principle" applies with SYMAP; the program normally stratifies the data into several classes or levels and applies a different pattern to each, and as few as four levels can be used to show all the polygons distinctly in mapping Thiessen polygons with SYMAP (Fig. 4). To do this one employs a set of dummy input data such that no two adjacent polygons will have the same level.

3. THIESSEN POLYGON COMPUTATIONS FOR SASKATCHEWAN CROP DISTRICTS

a. Determining the Thiessen weighting factors

The crop district boundaries (Fig. 1) had been digitized previously. Three reference points for which the latitudes and longitudes could readily be determined were selected from the digitized boundaries, and for these points the coordinates in the standard reference system were computed from the latitudes and longitudes. The digitized coordinates (XRD, YRD), were (24627, 887), (12198, 967) and (12955, 11361), and the standard coordinates (XRS, YRS), were (17.35, 7.29), (12.44, 8.14) and (13.41, 12.21) for the three reference points. The digitized coordinates are in the units used by the digitizing equipment, while the standard coordinates are in the units employed in the SYMAP program, but the methods described here would apply equally well in any other convenient system of measurement.

Eq. (2) and (3) were solved to obtain the A(I,J)'s and B(I)'s using the Gaussian elimination routine, and (1) was then applied to transforming the boundary coordinates to the standard system. For example, using the computed values of A(1,1), A(1,2), and B(1), for boundary point (5123,19662) the new X-value is $XS = .0003955(5123) + .00006452(19662) + 7.554 = 10.85$.

The standard coordinates for the stations were determined from the latitudes and longitudes. For example for station 14, Saskatoon (52°10' N, 106°41' W), use of the appropriate computer subroutine yielded the standard coordinates (XS, YS) = (14.81, 10.45) (Table 1).

The grid mesh program was then applied to the station and boundary standard coordinates to determine the Thiessen weighting factors. Examples are given here (Table 2) for

TABLE 1. Climatological stations with their numbers, latitudes, longitudes, and X-Y coordinates derived for a Lambert conformal conic projection.

Station	Number	Latitude	Longitude	X	Y
Lethbridge	1	49°38'	112°48'	11.03	9.09
Medicine Hat	2	50°1'	110°43'	12.25	9.10
Brooks	3	50°35'	111°54'	11.74	9.75
Calgary	4	51°6'	114°1'	10.75	10.50
Red Deer	5	52°11'	113°54'	11.08	11.38
Coronation	6	52°6'	111°27'	12.32	10.96
Vermilion	7	53°21'	110°50'	12.90	11.92
Edmonton	8	53°19'	113°35'	11.53	12.27
Cold Lake	9	54°25'	110°17'	13.40	12.75
Meadow Lake	10	54°7'	108°26'	14.25	12.29
Nipawin	11	53°20'	104°0'	16.36	11.22
Prince Albert	12	53°13'	105°41'	15.49	11.26
North Battleford	13	52°46'	108°15'	14.10	11.13
Saskatoon	14	52°10'	106°41'	14.81	10.45
Hudson Bay	15	52°52'	102°24'	17.14	10.72
Wynyard	16	51°46'	104°12'	16.07	9.90
Yorkton	17	51°16'	102°28'	16.94	9.34
Kindersley	18	51°28'	109°9'	13.39	10.13
Swift Current	19	50°17'	107°41'	13.96	8.95
Moose Jaw	20	50°20'	105°33'	15.14	8.78
Regina	21	50°26'	104°40'	15.64	8.79
Broadview	22	50°23'	102°35'	16.78	8.59
Rock Glen	23	49°10'	105°58'	14.74	7.81
Estevan	24	49°4'	103°0'	16.41	7.47
Brandon	25	49°52'	99°58'	18.21	8.04
Pilot Mound	26	49°12'	98°53'	18.77	7.37
Portage La Prairie	27	49°54'	98°16'	19.16	7.96
Winnipeg	28	49°54'	97°14'	19.74	7.94
Gimli	29	50°37'	96°59'	19.90	8.56
Dauphin	30	51°6'	100°3'	18.24	9.07
The Pas	31	53°58'	101°6'	17.90	11.59
Lacombe	32	52°28'	113°45'	11.23	11.60
Indian Head	33	50°32'	103°40'	16.20	8.79
Scott	34	52°22'	108°50'	13.72	10.85
Melfort	35	52°49'	104°36'	15.99	10.83
Morden	36	49°11'	98°5'	19.23	7.33

TABLE 2. Sample results for two crop districts of the iterative calculation of Thiessen weighting factors

District	Iteration	Total Number of Grid Points	Grid Points Within Boundary	Station Numbers with Factors Expressed as Percentage of Area							
4a	1	228	74	2	59.5	19	40.5				
	2	833	295	2	59.7	19	40.3				
	3	3,122	1,182	2	58.7	19	41.3				
	4	12,019	4,734	2	58.6	19	41.4				
2b	1	96	29	20	31.0	21	34.5	22	0	33	34.5
	2	327	131	20	30.5	21	36.6	22	0	33	32.8
	3	1,167	531	20	29.8	21	35.8	22	0	33	34.5
	4	4,406	2,126	20	30.3	21	36.0	22	0	33	33.7
	5	17,042	8,537	20	30.7	21	36.0	22	0.1	33	33.2
	6	66,791	34,179	20	30.7	21	36.2	22	0.1	33	33.1

TABLE 3. Thiessen weighting factors and stations for Saskatchewan crop districts.

District								
1a	24	.8856	22	.0944	25	.0200		
1b	22	.9564	33	.0405	25	.0030		
2a	21	.3738	24	.3518	33	.1884		
	20	.0628	22	.0167	23	.0065		
2b	21	.3622	33	.3306	20	.3067	22	.0006
3as	23	.7810	24	.1627	20	.0434	21	.0129
3an	20	.7931	19	.1810	23	.0259		
3bs	19	.7443	23	.2557				
3bn	19	.8991	18	.0790	14	.0219		
4a	2	.5858	19	.4142				
4b	18	.3744	19	.3252	2	.3004		
5a	17	.4351	33	.2486	22	.1842		
	16	.1013	21	.0303				
5b	16	.5133	17	.4072	15	.0562	35	.0234
6a	16	.3784	14	.2598	21	.1896	20	.1721
6b	14	.8806	20	.0905	13	.0145		
	18	.0084	12	.0053	19	.0007		
7a	18	.9453	34	.0324	14	.0223		
7b	34	.8111	13	.1146	18	.0551	14	.0193
8a	15	.3622	11	.3245	35	.3123	16	.0010
8b	35	.4921	14	.2365	12	.2346	16	.0369
9a	12	.4087	13	.2740	10	.1293		
	11	.1159	14	.0663	35	.0056		
9b	10	.3961	13	.2423	7	.1403	9	.1362
	34	.0851						

TABLE 4. June 1978 rainfall for all stations, and for Saskatchewan districts computed from the station data using the weighting factors from Table 3.

Station	Rain in mm	District	Rain in mm
1	17.6	1a	111.2
2	53.4		
3	60.7	1b	34.8
4	59.6		
5	112.6	2a	80.1
6	19.3		
7	39.5	2b	56.4
8	40.6		
9	15.6	3as	84.6
10	31.7		
11	115.2	3an	50.6
12	124.6		
13	45.7	3bs	64.5
14	39.7		
15	56.3	3bn	56.1
16	68.2		
17	55.0	4a	55.9
18	22.6		
19	59.4	4b	43.8
20	47.7		
21	56.8	5a	54.7
22	33.6		
23	79.4	5b	63.1
24	121.3		
25	28.7	6a	55.1
26	81.9		
27	40.5	6b	40.8
28	48.6		
29	49.0	7a	24.0
30	118.8		
31	40.4	7b	52.0
32	62.9		
33	64.2	8a	91.8
34	55.2		
35	108.8	8b	94.7
36	11.6		
		9a	84.2
		9b	36.0

two districts. For Saskatchewan District 4a, it was clear from the first that only stations 2 and 19 were involved, and by the fourth iteration the results had converged and indicated that 58.6% of the district area was closest to station 2 and 41.4% to station 19. For District 2b, at first only stations 20, 21 and 33 were involved but after the fourth iteration station 22 was also included, although only a very small part of the area was closest to that station.

To provide a visual representation, the SYMAP program was used to map polygons showing the areas closest to the various stations (Fig. 3). A crop district map was then overlaid on this and for each district, the portion of the area of each polygon that fell within the district was inspected and the corresponding final weighting factor from the grid mesh calculations (Table 3) was checked for reasonableness. This visual inspection also helped us to assess the adequacy of spatial representation. For example District 2b, with three stations well distributed within the district, is quite adequately represented for purposes such as large area yield analysis. In contrast District 4a is covered by the polygons of two stations some distance outside the district, to the northwest and northeast, and is therefore not well represented.

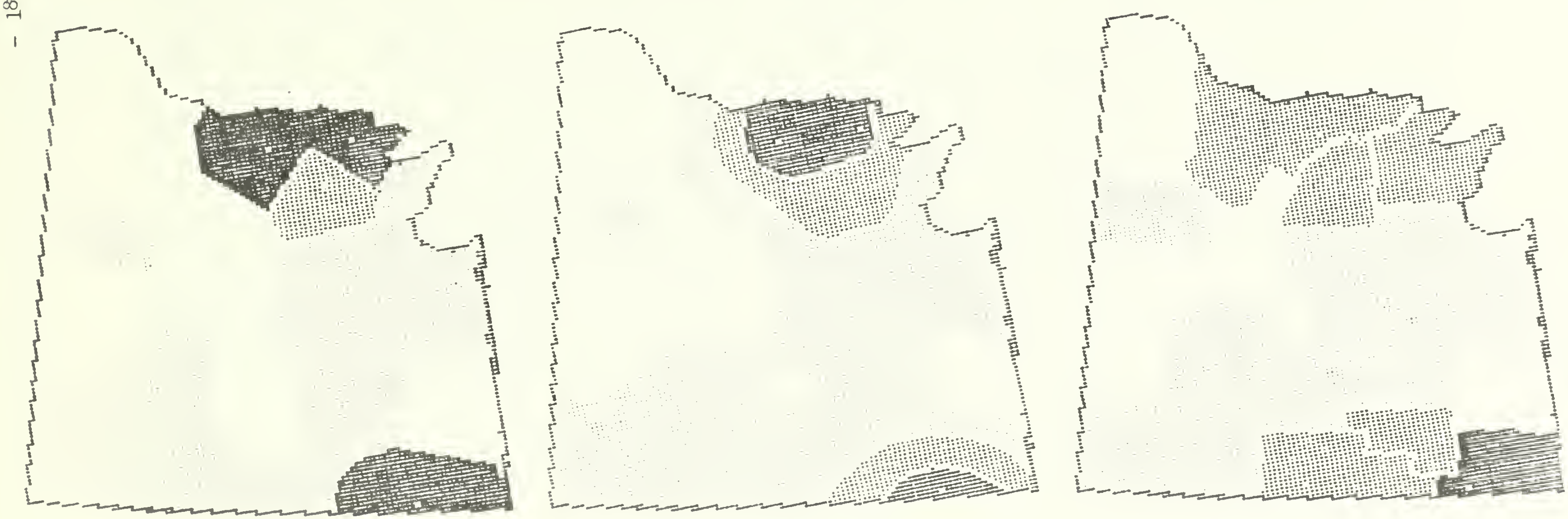
<u>b</u>. <u>Applying the weighting factors to computing district rainfall</u>

The application of the weighting factors is here illustrated by an example taken from the 1978 yield forecast computations carried out in Agriculture Canada. These computations included the transformation of rainfall data from a station to a district basis (Table 4). For example, the June 1978 District 4a rainfall was computed from the amounts for stations 2 and 19 as: $.586(53.4) + .414(59.4) = 55.9$ mm.

A visual representation of the station data on a polygon basis (Fig. 5a) appears quite different from that of the same data using isolines (Fig. 5b). The isolines give a more realistic visual representation of the spatial patterns of the rainfall, but for the purpose of computation of district rainfall, it is much more convenient to use the polygon approach.

Applying the Thiessen weighting factors in computing the district values in effect spatially smoothes the data, so the district data tend to be less variable than are the station data on which they are based, even when the number of districts is relatively large. For example, in the northern part of the Saskatchewan agricultural region, of the four rainfall classes only the lowest three appear in the district data (Fig. 5c), but in the mapping based on station data all four of the levels, incuding the extreme highest (darkest shading) appear in the

Fig. 5 June 1978 rainfall (Table 4) mapped with SYMAP. White indicates less than 50 mm, lightest shading: 50 to 80, next lightest: 80 to 110, darkest: over 100 mm of rain. _a_. (left) Station data on Thiessen polygon basis. _b_. (centre) Station data mapped to show isohyets rather than shaded polygons. _c_. (right) Calculated district rainfall.



northern area. (Figures 5a and b). The sample standard deviation for the 20 Saskatchewan districts is 22.7 mm, while for the 18 Saskatchewan stations it is 31.5.

4. DISCUSSION OF AREAL DISTORTION

The Lambert Conformal Conic projection, which was used here, has several advantages for this work, but as with all conformal projections it exaggerates or reduces relative areas from one part to another of the map (Robinson and Sale, 1969). It could be suggested that the use of the Thiessen polygon with this projection could lead to errors because equal areas on the earth's surface often do not correspond to equal areas on the map. To investigate the extent of this problem for the region considered, it was decided to calculate and compare projected areas in the north and the south of Saskatchewan. Triangles were used because of the ease with which areas can be evaluated given the coordinates of their vertices either in spherical or rectangular coordinates.

The northern triangle was enclosed by the lines joining stations 10, 12 and 13, and the southern triangle was formed by lines joining 19, 20 and 23 (Fig. 1). The latitudes, longitudes and the X-Y coordinates of these points are indicated in Table 1. Using latitudes and longitudes and spherical trigonometry methods (Meyler and Sutton, 1958), the calculated area of the northern triangle was 13242 km^2 and that of the southern triangle was 10011 km^2. Using the rectangular coordinates for the triangle vertices presented in Table 1 and standard methodologies for determining triangle areas, the projected area of the northern triangle was found to be 0.79645 $units^2$ and the projected area of the southern triangle was 0.6063 $units^2$. Reducing these results to comparable terms, it is found that 1.0 km^2 in the northern triangle corresponds to 6.015×10^{-5} $units^2$ in the projection and 1.0 km^2 in the southern triangle corresponds to 6.056×10^{-5} $units^2$ in the projection. A km^2 therefore is projected on a smaller area in the north than in the south. The northern area in the projection representing one km^2 is equal to 99.3% of the area representing one km^2 in the south.

5. CONCLUSIONS

This paper describes and demonstrates a practical system for employing readily available point and boundary location information to derive weighting factors for use in computing areal meteorological data from station data. It is particularly effective for applications, such as large area yield prediction, where a number of bounded districts is involved and there are frequent changes in the station set. The system includes computer mapping procedures based on SYMAP, to aid the user in understanding the spatial relationships and in assessing the suitability of the station distribution.

The study confirms the effectiveness of the grid mesh technique for calculating the Thiessen weighting factors, especially in comparison with traditional manual techniques. The computer implementation of this grid mesh technique proved to be very efficient for the sample application considered here. The computer time requirements depend largely on the level of accuracy required and the resulting number of iterations.

In order to use the grid mesh technique, the station and district boundary coordinates must be available in a compatible form. The system presented here demonstrates how quite standard source data can be reduced to the required form. The linear transformation was shown to be effective in reducing large amounts of data from different X-Y coordinates system to a common base.

Usually the digitized boundary coordinate set should include at least three reference points. Where a previously digitized set that lacks such reference point coordinates is being employed, the ability of the authors' system to specify the transformation using only two reference points can considerably reduce the effort required.

The use of the Lambert conformal conic projection with the specified standard parallels was found to be acceptable for area determinations in the region of interest.

The source map and computer maps can be any convenient scale, and the system could readily be adapted for use with other map projections for which the computer routines for converting from latitude and longitude were available. For example the Universal Transverse Mercator (UTM) projection, which is appropriate for detailed studies of relatively small areas and is commonly used in topographic maps, provides a rectangular grid system. These UTM grid coordinates can be input for SYMAP mapping (Williams, McKenzie and Sheppard, 1980). They could also be used directly in the Thiessen grid mesh program, and the linear transformation described here could readily be employed to reduce boundary coordinates digitized on a UTM map to the UTM grid system.

The system developed by the authors could be employed in cartographic data structure analyses. The districts themselves are polygons, so the situation (Fig. 4) is one of overlapping polygons. Peucker and Chrisman (1975) suggest that in such cases attention be directed to the "Least Common Geographic Units" (LCGU's), i.e. the areal units uncut by any lines. These LCGU's are here the basis of the Thiessen weighting factors. For example, there are two LCGU's in District 4a, and the ratio of their areas, 5858:4142, indicates the relative weights to be given to data from stations 2 and 19 in computing areal values for 4a (Fig. 4 and Table 3).

Acknowlegements

The required computer programming was done in the Agrometeorology Section, and computer facilities were provided by the Data Processing Division of Agriculture Canada. R. Muma of the Agrometeorology Section used SYMAP to prepare the computer maps. Computer subroutines for transforming from latitude and longitude to X-Y coordinates were obtained from the Applications Program Library of the Canada Department of Energy, Mines and Resources, and N.J. Pocock of that Department provided information on the reasons for using various map projections. The manuscript was reviewed by R.B. Stewart and P.Y.T. Louie.

REFERENCES

Dahlquist, G., A. Bjorck and N. Anderson, 1974: Numerical Methods. Prentice-Hall, Englewood Cliffs, N.J. 573 pp.

Dougenik, J.A. and D.E. Sheehan, 1976: SYMAP User's Reference Manual Fifth Edition. Harvard University, Cambridge, Mass. 200 pp.

Louie, P.Y.T., 1977: Thiessen Polygon Coefficients by a grid mesh technique. TEC 847, Environment Canada, Atmospheric Environment Service, Downsview, Ontario. 16 pp.

Meyler, D.S. and O.G. Sutton, 1958: A Compendium of Mathematics and Physics. English Universities Press, London. 384 pp.

Newton-Smith, W.H. and G.D.V. Williams, 1964: The relationship between rain gauge network density and areal rainfall estimates on the prairies. Meteorological Branch, Canada Dept. of Transport, Circular Tec-542. 17 pp.

Peucker, T.K. and N. Chrisman, 1975: Cartographic data structures. American Cartographer 2,55-69.

Robinson, A.H. and R.D. Sale, 1969: Elements of Cartography, Third Edition. John Wiley and Sons, New York, N.Y. 415 pp.

Rhynsburger, D., 1973: Analytic delineation of Thiessen polygons. Geographical Analysis. 5,133-144.

Sheppard, M.I. and G.D.V. Williams, 1976: Quantifying the effects of great soil groups on cereal yields in the Prairie Provinces. Can. J. Soil Sci. 56,511-516.

Starr, T.B. and P.I. Kostrow, 1978: The response of spring wheat yield to anomalous climate sequences in the United States. J. Appl. Meteor. 17, 1101-115.

Williams, G.D.V. and G.W. Robertson, 1965: Estimating most probable prairie wheat production from precipitation data. Can. J. Plant Sci. 45,34-47.

Williams, G.D.V. and W.R. Sharp, 1972: Computer mapping in agrometeorology. Tech. Bull. 80, Agrometeorology Section, Agriculture Canada, Ottawa. 40 pp.

Williams, G.D.V., M.I. Joynt and P.A. McCormick, 1975: Regression analyses of Canadian Prairie crop-district cereal yields 1961-1972, in relation to weather, soil and trend. Can. J. Soil Sci. 55,43-53.

Williams, G.D.V. and R.G. Lattimore, 1978. Weather-based prairie crop yield forecasts, 1978. Market Commentary, Grains and Oilseeds, July 1978. Policy, Planning and Evaluation Branch, Agriculture Canada, Ottawa. 23-25.

Williams, G.D.V., J.S. McKenzie and M.I. Sheppard, 1980: Mesoscale agroclimatic resource mapping by computer, an example for the Peace River region of Canada. Agricultural Meteorology 21, 93-109.

Williams, G. Daniel V.
Procedures for computing and mapping Thiessen weighting